

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局



(10) 国际公布号
WO 2024/235160 A1

(43) 国际公布日
2024 年 11 月 21 日 (21.11.2024)

(51) 国际专利分类号:
G06F 40/166 (2020.01)

(21) 国际申请号: PCT/CN2024/092636

(22) 国际申请日: 2024 年 5 月 11 日 (11.05.2024)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
202310538686.9 2023 年 5 月 12 日 (12.05.2023) CN

(71) 申请人: 北京京东尚科信息技术有限公司 (BEIJING JINGDONG SHANGKE INFORMATION TECHNOLOGY CO., LTD) [CN/CN]; 中国北京市海淀区知春路 76 号 8 层, Beijing 100086 (CN)。

(72) 发明人: 李浩然 (LI, Haoran); 中国北京市海淀区知春路 76 号 8 层, Beijing 100086 (CN)。吴友政 (WU, Youzheng); 中国北京市海淀区知春路 76 号 8 层, Beijing 100086 (CN)。

(74) 代理人: 北京品源专利代理有限公司 (BEYOND ATTORNEYS AT LAW); 中国北京市海淀区莲花池东路 39 号西金大厦 6 层, Beijing 100036 (CN)。

(81) 指定国(除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA,

(54) Title: TEXT GENERATION METHOD AND APPARATUS, DEVICE, AND STORAGE MEDIUM

(54) 发明名称: 一种文本生成方法、装置、设备和存储介质

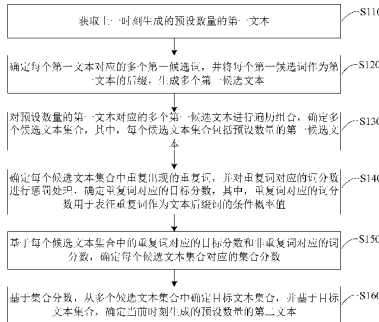


图 1

- S110 Acquire a preset number of first texts generated at a previous moment.
S120 Determine a plurality of first candidate words corresponding to each first text, and use each first candidate word as a suffix of the first text to generate a plurality of first candidate texts.
S130 Traverse and combine the plurality of first candidate texts corresponding to the preset number of first texts to determine a plurality of candidate text sets, wherein each candidate text set comprises a preset number of first candidate texts.
S140 Determine repeated words in each candidate text set, and perform penalty processing on word scores corresponding to the repeated words to determine target scores corresponding to the repeated words, wherein the word scores corresponding to the repeated words are used for representing conditional probability values of the repeated words being used as text suffix words.
S150 On the basis of the target scores corresponding to the repeated words in each candidate text set and word scores corresponding to non-repeated words, determine a set score corresponding to each candidate text set.
S160 On the basis of the set score, determine a target text set from among the plurality of candidate text sets, and, on the basis of the target text set, determine a preset number of second texts generated at a current moment.

(57) Abstract: Disclosed in embodiments of the present application are a text generation method and apparatus, a device, and a storage medium. The method comprises: acquiring first texts generated at a previous moment; determining a plurality of first candidate words corresponding to each first text, and using each first candidate word as a suffix of the first text to generate a plurality of first candidate texts; traversing and combining the plurality of first candidate texts corresponding to the preset number of first texts to determine a plurality of candidate text sets; on the basis of each word in each first candidate text, determining repeated words in each candidate text

[见续页]

PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW。

(84) 指定国(除另有指明，要求每一种可提供的地区保护): ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告(条约第21条(3))。

set, and performing penalty processing on word scores corresponding to the repeated words to determine target scores corresponding to the repeated words; on the basis of the target scores corresponding to the repeated words in each candidate text set and word scores corresponding to non-repeated words, determining a set score corresponding to each candidate text set; and determining a target text set on the basis of the set score, and, on the basis of the target text set, determining second texts generated at a current moment.

(57) 摘要: 本申请实施例公开了一种文本生成方法、装置、设备和存储介质。该方法包括: 获取上一时刻生成的第一文本; 确定每个第一文本对应的多个第一候选词, 并将每个第一候选词作为第一文本的后缀, 生成多个第一候选文本; 对预设数量的第一文本对应的多个第一候选文本进行遍历组合, 确定多个候选文本集合; 基于每个第一候选文本中的每个词, 确定每个候选文本集合中的重复词, 并对重复词对应的词分数进行惩罚处理, 确定重复词对应的目标分数; 基于每个候选文本集合中的重复词对应的目标分数和非重复词对应的词分数, 确定每个候选文本集合对应的集合分数; 基于集合分数确定目标文本集合, 并基于目标文本集合确定当前时刻生成的第二文本。